ENGLISH EDUCATION JOURNAL
(EEJ)
Journal Homepage: http://www.jurnal.unsyiah.ac.id/EEJ/

LEXICAL BUNDLES IN STUDENTS' ESSAY WRITING

Nadia Ulfa[1]
Kismullah Abdul Muthalib

Universitas Syiah Kuala, Banda Aceh, Indonesia

ABSTRACT

This study was aimed to find out the frequency of three- word and four-word of lexical bundles which were used in the essay writing of the undergraduate students studying at English Education Department of Syiah Kuala University. This study employed a corpus based analysis which compiled a specialized small corpus consisting of 161 students' essays under linguistic topic. In order to generate the frequent list of sequences, computational software named AntConc 3.5.8 by Laurent Anthony was used. The results showed that the number of three-word bundles was higher than four-word bundles in terms of the frequency of co-occurrence in which the most common expression *the function of* as three-word bundles and *the advantages of being* as four-word bundles were found.

Keywords: *word combinations, lexical bundles, corpus based analysis.*

INTRODUCTION

University students generally deal with academic assignments, for example writing essays, theses, and articles for journals. In this case, many students face obstacles in using both language use and rules that are required in an academic writing. A study which was carried out by Pratiwi (2016) found that the undergraduate students of English Education Program in University of Bengkulu mostly dealt with some difficulties in writing. The study which involved the assessment of the students' writing showed that they tended to face linguistics difficulties such as language use and vocabulary aspects much more than content aspect (Pratiwi, 2016). It should be noted that language use is one of the components that students need to make a good writing. As Heaton (1975)

[1] Corresponding author: nadiaulfa.nafa@gmail.com

stated, language use relates to the ability to use appropriate sentences and to write them correctly. Therefore, when university students had knowledge about how expert writers usually arrange words in sequences in an academic writing, their writing may improve.

It is important for university students to recognize experts' writing by studying any corpus which relates to their disciplines. This is aimed to show them about how to a piece of academic academically. In this case, the knowledge of word combinations is essential as it contributes to a coherent discourse. It should be noted that one of the word combinations which give influences to writing is lexical bundles.

LITERATURE REVIEW

Lexical Bundles

Lexical bundles are widely known by other terms such as "phraseology" (Meunier & Granger, 2007), "formulaic language" (Schmitt & Carter, 2004). Moreover, Prior studies which have noted the emergence of lexical bundles were once initiated in 1924 by Jespersen and Firth in 1951 when the terms "collocation" and "fixed expression" turned up (Kashiha & Chan, 2013, p. 134). These studies are considered to be the "pioneer" of such word combinations study. The term lexical "bundles itself" was popularized by Biber, Johannson, Leech, Conrad and Finegan in 1999. Biber et al. (1999, p. 990) defined lexical bundles as "Recurrent expression, regardless of their idiomaticity, and regardless of their structural status".

In order to be considered as lexical bundles, there are certain criteria that must be met. One of those has been stated by Biber et al. (1999) that the word combinations should show a statistical tendency to occur. Furthermore, Biber, Conrad and Cortes (2004) assumed in their study on lexical bundles that frequency has the key role in identifying lexical bundles in a corpus. The frequency identification of lexical bundles varied since the cutoff point of frequency was different in various studies. This criterion should meet the specified cutoff points that differ, depending on the size of the corpus used in the study. As in this study, because of using a smaller corpus, the cutoff point used was adopted from Biber and Barbieri (2007) which stated that for the smaller sub-corpus size of the study which range to fewer than 40.000 words, lexical bundles are defined as words combination that occur three times or more

at least in three texts.

Basically, word combinations can be clustered into two to six word strings in the corpus software. However, common word clusters found in any ready corpora is three-word string. This happens because two word strings is often considered as collocations. For the research purpose, the set of word clusters can be adjusted by the researcher through the software, because the string itself depends on the size of corpus which will be analyzed. For consideration, this study generated three-word and four-word strings to be analyzed in order to find out more about this issue.

The Role of Lexical Bundles in Academic Writing

Lexical bundles have an important role that become characteristic of certain registers. A study conducted by Biber et al. (1999) found that there were characteristics of using certain lexical bundles in each register such as conversation and academic prose. In their research, it was mentioned that I don't know (p. 994) with 1,000 occurrences per million words was found to be the most common expression in conversation. Other examples of lexical bundles found in conversation was I don't think and I don't want (p. 994) with each occurring 400 and 200 times per million of words. Meanwhile, in academic prose, the expressions such as in order to, one of the, the number of (p. 994) became the most frequently used with over 200 occurrences. It can be inferred that the use of lexical bundles is not similar even in many disciplines. From the academic point of view, knowledge of lexical bundles is very important as Coxhead and Byrd (2007) claimed that one of the importance roles of lexical bundles is to assist in defining a fluent use of a language. Furthermore, Biber et al. (2004) argued that students should pay attention to word combinations which are typically used in their given discipline as they will assist them to understand the word combinations frequently found in the discipline.

Numerous researches in Indonesia have given great attention to the importance of lexical bundles in academic writing, including those conducted by Samodra and Pratiwi (2018). In their research, they compared the use of lexical bundles between Indonesian and English abstract theses. The study involved 140 theses abstract written by undergraduate students from two disciplines, namely social and natural sciences. The results of this study indicated that the lexical bundles "in this research*"* was mostly found. Many studies related to lexical bundles show that the awareness and knowledge of the use of lexical bundles are

important to consider.

Coxhead and Byrd (2007) proposed three main reasons about why lexical bundles are important for writers. First, the repetition of lexical bundles can construct language and provide ideas to both speakers and writers in building discourse. Second, the use of lexical bundles accounts for fluency in a language and holds a certain accreditation and legitimacy in the discipline one writes or speaks in. Third, the classification of lexical bundles exposes their grammatical structure and leads to their function in language, such an indication for writers or speakers to use proper word combination in an appropriate context.

In academic writing context, the knowledge about word combinations may be crucial for university students to know in order to notice how specific word combinations used are related to their discipline. Heaton (1989) listed some aspects in academic writing; they are mechanics, sentences, vocabulary, unity, cohesion, coherence, organization, and topic. Among the aspects which were proposed by Heaton (1989), cohesion and coherence are two of the aspects which contribute much to a fluent academic writing. Generally, cohesion and coherence are two terms which are related to making sense of language in a text (Poudel & Dhankuta 2018). Another important thing that should be considered in an academic writing is getting the academic style of writing. The academic style of writing encompasses a clear and precise formal grammar and vocabulary, maintains the cohesive ideas, and builds coherence in a text (Poudel & Dhankuta 2018). Those all characteristics are mostly found in experts' writing in any academic text collection in the form of books or articles.

It should be noted that there are various disciplines used mostly different word combinations in terms of their specific terms. Through corpus analysis, university students can observe frequent word combinations used in expert studies which further help them to be fluent in writing. As Tribble (2002) assumed, by studying corpus, learners can draw generalization on specific genres that they are studying which in the end can assist them in writing. Among the writing genres, there are those that require students to write in a formal context when formal words and grammar are preferred to use. Furthermore, formal words and grammar used in that formal context are related to the procedure for using the word combination pattern in English. Therefore, the students' awareness in combining words in writing is crucial since an academic

writing probably becomes one of the challenging studies for university students because it needs some complex requirements from lexical choice, word combinations to the structure requirement in any disciplines.

By considering the role of lexical bundles, Biber et al. (2004) stated that the awareness of the presence and function of lexical bundles in academic English is crucial to students and teachers because the presence and the function of lexical bundles can lead to a deeper understanding of the discourse construction. Furthermore, the knowledge about lexical bundles which the students have will not be limited to their final requirement in university, but also to help them in developing their writing competence either in essays, theses or articles.

RESEARCH METHODOLOGY

Research Subject

The subject of this study was a specialized corpus which compiled 161 essays with 69.834 words in the field of linguistic. Those essays were written around 2018 and 2019 as a part of Introduction to Linguistics assignments by the students of English Education Department in Syiah Kuala University. Meanwhile, the object of this study was the word combinations generated from AntConc software developed by Laurent Anthony.

Research Instrument

Documentation

Documentation is used in order to collect the data of this study. In documentation analysis research, the documentation in the form of students' essay writing was used to be analyzed. The essays which were used in this study varied in topics in the Linguistics field. Furthermore, the essay collection used in this study was unedited essays where the researchers did not change anything from the essay contents.

Corpus Software

There are various computational devices which can be utilized in corpus linguistics study. However, Computer software named Antconc 3.5.8 for Windows (which was downloaded from http://www.laurenceanthony.net) became an instrument used in this study to discover the word patterns. It should be noted that "AntConc is a freeware, multiplatform tool for carrying out corpus linguistics

research and data-driven learning" (Anthony, 2019).

Technique of Data collection

The data collection technique used in this study was documentation. A collection of essay writing of the English Department students in Syiah Kuala University which was obtained from the English Education Department lecturers. Further, this study employed the steps of corpus data collection from Biber (2006, pp. 23-31) as follows:

1. Designing and constructing a corpus.

This study designed a specialized corpus in the field of academic writing which consisted of students' essay writing. The length of the students' essay was about 200 to 300 words each essay. It should be noted that the students' essays which were used in this study varied in topics.

2. Transcribing and scanning the data.

Collecting the data in this study focused on gathering the written texts. The written texts then were partially scanned or retyped manually into *docx* format. Meanwhile, in the next process it involved the steps of digitally transcribing the essays into *text* format due to the software demand (the standard format used in a word-processing program is *text* format). The information related to author names, figures, numbers, tables and references was excluded before extracting it into the computer software.

3. Analyzing a corpus.

In this step, the use of AntConc seemingly became a crucial thing where the students' essays in text format (.txt) were extracted to Antconc 3.5.8 software in order to retrieve the word clusters.

Technique of data analysis

Adapted from Salazar (2011, pp. 47-53), there were some stages in analyzing lexical bundles. One of the stages was lexical bundles identification. As the first step in the analysis, lexical bundles identification was intended to create the list of the most frequent lexical bundles found in the English Department students' essay writing. In order to generate and arrange the list, AntConc 3.5.8 was utilized. It should be noted that only word combinations that meet the criteria were analyzed in this study.

RESULTS AND DISCUSSIONS

Below is the chart of the frequency of lexical bundle occurrences between three-word and four-word bundles found in the corpus using AntConc software.

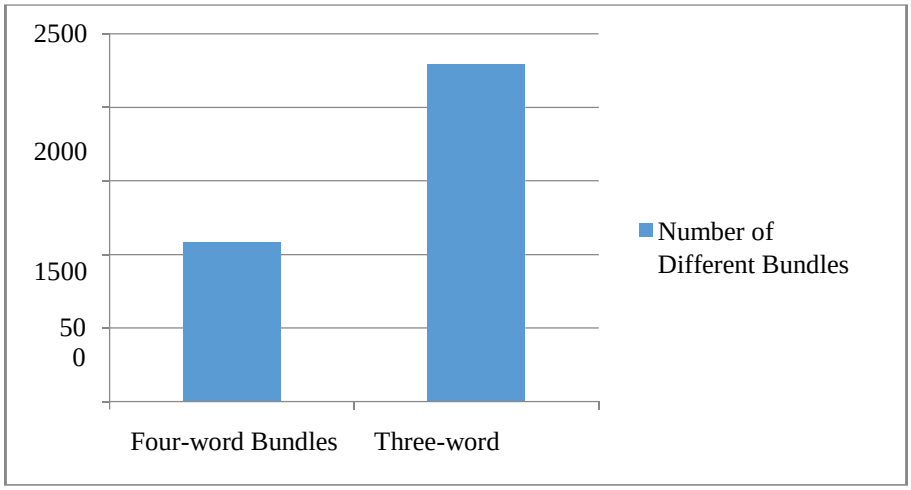

Figure 1. The Number of three-word and four-word bundles found in the corpus

Data above revealed that the number of occurrences between three-word and four-word bundles was different. The chart shows that there are more than 1000 bundles (different words) of four-word clusters in 5337 tokens (total number of words). On the other hand, three-word bundles have number occurrences reaching 2000 bundles types in 15295 tokens.

However, in this case, not all bundle types were analyzed because some do not meet the criteria set by Biber et al. (1999). In order to filter the criteria, the use of exclusion criteria was necessary to exclude the sequences. The use of exclusion criteria was important to more easily manage since AntConc produced such a long list of lexical bundles to be analyzed as showed by the total number in Figure 1 above. The exclusion criteria which were used in this study was adopted from Salazar (2011) which consisted of topic-specific bundles, bundles which consisted of random numbers, temperature, volume, and length bundles, random section bundles, and meaningless bundles.

Below is the list of lexical bundles generated from AntConc after applying the exclusion criteria (see Table 1). The list was arranged from

the most to the least frequent lexical bundles found in the students' essay corpus.

Table 1. The Top 20 Three-Gram Lexical Bundles

No.	List of Three-grams bundles	Frequency	Range
1	The function of	72	40
2	In the future	53	41
3	Be able to	42	32
4	The varieties of	42	25
5	Advantages of being	3	35
6	is very important	34	25
7	as long as	31	24
8	in the world	31	21
9	One of the	31	21
10	The advantages of	31	30
11	Around the world	29	22
12	Is not a	29	24
13	A lot of	28	21
14	Be aware of	27	21
15	The use of	26	19
16	because of the	25	21
17	In my opinion	25	21
18	In order to	25	20
19	To be a	25	18

20	To produce the	24	18

As can be seen in Table 1, the most notable pattern of three- word bundles found in the corpus was noun phrase with embedded *of* fragment *the function of* (found 72 times in 40 essays). Following this, the construction of prepositional phrase took the second place in terms of the high frequency, *in the future*, which occurred 53 times in 41 different essays, and followed by *be able to* and *the varieties of* with 42 occurrences each.

Meanwhile, the results showed by AntConc on the four- word bundles have different sets of expressions. This can be seen from Table 2 as follows.

Table 2. The top 20 Four- Gram Lexical Bundles

No.	List of Four-grams bundles	Frequency	Range
1	the advantages of being	28	27
2	advantages of being a	25	24
3	be aware of the	23	20
4	aware of the varieties	22	20
5	of the varieties of	20	18
6	i would like to	14	11
7	to be aware of	13	13
8	Is one of the	11	10
9	It is important for	11	10
10	Should be aware of	11	7
11	To be able to	11	11
12	I think it is	10	5
13	in the future because	10	10
14	in the process of	10	8

15	help me become a	9	9
16	is not a problem	9	8
17	it is not a	9	9
18	will be able to	9	8
19	is the study of	8	6
20	is very important to	8	6

Regarding to four-word bundles, Table 2 shows that the frequency of pattern co-occurrences consists of four-word bundles were in a small amount found in the corpus. As can be seen, the construction in the form of noun phrase with embedded *of* fragment has the highest occurrence in the corpus (e.g. *the advantages of being*, *advantages of being a*, respectively occurred 28 and 25 times). It was followed by the expression *be aware of the* which is in the form of copula *be* + noun phrase/ adjective phrase with 23 frequency of co-occurrences.

Discussions

Based on the result described in the previous section, the sequences of lexical bundles vary depending on the type and size of corpus and register of the study. In addition, the determination of cutoff points in categorizing lexical bundles also affected the results of word combinations generated in the software. As showed in the result, the number of lexical bundles was found in large quantities due to the use of small cutoff points as mentioned by Biber and Barbieri (2007) that for smaller sub-corpus about 50.000 word, the word combinations must occur in 3 different texts to be considered as lexical bundles. Consequently, as stated in the result section, in terms of three-word bundles, there were 2648 bundle types in 15.295 tokens. Other than that, in terms of four-word bundles, there were 1081 bundle types in 5337 tokens. This is in line with what has been stated by Ucar (2017) that the string consisting of three-word lexical bundles are more commonly used in academic writing than longer lexical bundles.

Furthermore, as showed in the result, when lexical bundles are arranged based on frequency of co-occurrence after the exclusion criteria

adopted from Salazar (2011) were applied, the highest occurrence of three-word bundles was *the function of* (72 occurrences in 40 essay in the corpus). Other constructions which were identified in large quantities were *in the future* (other prepositional phrase construction) and *be able to* (verb/ adjective +) clause fragments. Meanwhile, in the form of four-word bundles, the expression *the advantages of being* was found as the highest occurrence. It is different from the results of the research conducted by Samodra and Pratiwi (2018) which indicated that other prepositional phrase fragment took the first place in terms of its frequency. The research-oriented expression, such as *in this research* was mostly found in the corpus. This could happen due to the translation and use of different research subject where they used abstracts from undergraduate thesis as their research. Conversely, this study used students' essays as the subject.

CONCLUSION AND SUGGESTION

As regard to the explanation in the previous sections, it could be concluded that each register, a conversation or written, has special characteristics in the use of such combinations. The corpus which consists of 161 academic essays and more than 60 thousand words about linguistic has produced thousands of bundles which were then filtered according to exclusion criteria. It should be noted that frequency of occurrence was the main focus that must be known in this study. Through frequency, the tendency of a corpus in the use of lexical bundles was identified. The findings show that the number of three word bundles was more than the four word bundles. After the exclusion criteria were applied to the list bundles generated by AntConc, it was identified that the first sequence with the highest frequency in this corpus were *the function of* in terms of three-word bundles and *the advantages of being* in terms of four-word bundles. Both were in the form of noun phrase followed by *of* fragment.

Further, it is crucial to note that students should be able to distinguish the use of clusters of lexical bundles according to their respective goals. In this case, lecturers must also be aware of this issue by involving corpora in their teaching and involving a corpus based learning so that their learners can distinguish the use of certain clusters that become the characteristic of the corpora used.

One should note that this study is expected to provide a useful reference for future research by using a larger corpus size and broad

scope since this study only examined the three and four word bundles which were found in a smaller corpus. Further, because this study only limits the corpus related to linguistics, it is expected that further research can examine many other fields so as to add insight into lexical bundles used in different fields of study.

REFERENCES

Anthony, L. (2019). *AntConc (Version 3.5.8) [Computer Software]*. Tokyo, Japan: Waseda University. Retrieved February 2, 2019 from https://www.laurenceanthony.net/software.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers.* Amsterdam, Philadelphia: John Benjamins.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263–286.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371– 405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing, 16*(2007), 129-147.

Heaton, J. B. (1975). *Writing English language tests*. London: Longman.

Heaton, J. B. (1989). *Writing English language test*. New York, NY: Longman.

Kashiha, H., & Chan, S. H. (2013). An exploration of lexical bundles in academic lectures: Examples from hard and soft sciences. *Journal of Asia TEFL, 10*(4), 133-161.

Meunier, F., & Granger, S. (2007). *Phraseology in foreign language learning and teaching.* Amsterdam: John Benjamins Publishing.

Poudel, A. P., & Dhankuta, N. (2018). Academic writing: Coherence and cohesion in paragraph. Retrieved August 8, 2019 from https://www.academia.edu/35687243/Academic_writing_Coherence_and_Cohesion_in_Paragraph

Pratiwi, K. D. (2016). Students' difficulties in writing English (A study

at the third semester students of English education program at University of Bengkulu Academic Year 2011-2012). *Linguists: Journal of Linguistics and Language Teaching*, *3*(1).

Salazar, D. J. L. (2011). *Lexical bundles in scientific English: A corpus-based study of native and non-native writing* (Doctoral Dissertation). Universitat de Barcelona, Barcelona. Retrieved February 15, from http://www.tdx.cat/bitstream/handle/10803/52083/DJLS_DISSERTATION.pdf?se quence=1.

Samodra, C. S., & Pratiwi, R. D. V. (2018). Lexical bundles in Indonesian and English thesis abstracts. *Proceedings of the Fourth Prasasti International Seminar on Linguistics. Advances in Social Science, Education, and Humanities Research, 166,* 183–191. Atlantis Press. https://dx.doi.org/10.2991/prasasti- 18.2018.35.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic and sequences: Acquisition, processing use* (pp. 1–22). Philadelphia: John Benjamins.

Tribble, C. (2002). Corpora and corpus analysis. In J. Flowerdew (Ed.), *Academic discourse* (pp. 131-149). London: Longman, Pearson Education.

Ucar, S. (2017). A corpus-based study on the use of three-word lexical bundles in the academic writing by native English and Turkish non-native writers. *English Language Teaching*, *10*(12), 28- 36.